Vision Research 48 (2008) 2523–2536

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

A critical band of phase alignment for discrimination but not recognition of human faces

Bruce C. Hansen [a,b,*], Reza Farivar [c], Benjamin Thompson [b,1], Robert F. Hess [b]

[a] Department of Psychology, Colgate University, 13 Oak Drive, Hamilton, NY 13346, USA
[b] McGill Vision Research Unit, Department of Ophthalmology, McGill University, Montreal, Que., Canada H3A 1A1
[c] Department of Psychology, McGill University, Montreal, Que., Canada H3A 1A1

ARTICLE INFO

Article history:
Received 9 April 2008
Received in revised form 18 August 2008

Keywords:
Face discrimination
Face recognition
Spatial phase alignment
Fourier filtering
Spatial frequency
Face frequency

ABSTRACT

We investigated the processes underlying the discrimination and recognition of human faces as a function of spatial phase alignment to assess whether face processing can be understood in terms of the amplitude spectrum alone. Specifically, we varied the amount of aligned Fourier phase in different regions of the face frequency spectrum and argue that the properties of the underlying neural processes are best understood in terms of the number of phase alignments as opposed to octave bandwidths. Additionally, we observed performance differences for face discrimination tasks compared to face recognition tasks. For face recognition, our results show that a narrower range of phase alignment is needed for face frequencies near 9 cpf when compared to 3 and 27 cpf, thereby supporting the notion of a critical frequency for face recognition. However, for face discrimination where participants were required to discriminate between an average face and different unique faces along a face morph continuum, performance depended on a fixed signal-to-noise ratio of phase alignment within a contiguous range of face frequencies (termed critical band of phase alignments), regardless of the central face frequency of that range within the face frequency spectrum when compared to non-phase randomized control thresholds.

1. General introduction

The early processing of visual content involves a spatial decomposition of the image by cells in the primary visual cortex that have band-pass properties for spatial frequency and orientation (e.g., DeValois, Albrecht, & Thorell, 1982; Maffei & Fiorentini, 1973). Psychophysically, the notion of spatial frequency "channels", introduced some 40 years ago by Campbell and Robson (1968) is thought to represent the properties of individual cells at this early stage of the visual pathway. Such channels are thought to be approximately 1 octave in bandwidth (but see Wilson, McFarlane, & Phillips, 1983) and have an orientation bandwidth of approximately 10–20° (Blakemore & Campbell, 1969; Phillips & Wilson, 1984; Wilson & Bergen, 1979). The usefulness of this channel concept has extended beyond its initial application to the detectability of relatively simple (in the Fourier sense) localized stimuli, to spatially complex objects in motion (Anderson & Burr, 1985; Hess, Bex, Fredericksen, Brady, 1998; Ledgeway, 1996; but also see Hess, Wang, & Liu, 2006), stereo (Heckmann & Schor 1989; but also see Hess et al., 2006), and to object recognition in general (Braje, Tjan, & Legge, 1995). On the other hand, there is relatively good evidence for the rigid *combination* of channel-based content in the detection of the motion and disparity of spatially complex objects (Hess et al., 2006) as well as the discrimination of spatially complex objects (Olzak & Wickens, 1997).

With respect to object recognition, it has been argued that letter identification can be understood in terms of a single elementary spatial channel with a bandwidth of 1.6 ± 0.7 (Majaj, Pelli, Kurshan, & Palomares, 2002; Solomon & Pelli, 1994). A similar claim has been made for face recognition, but for a central face frequency of approximately 10 cycles/face (Näsänen, 1999). In fact, a number of studies, using a variety of different techniques, have argued for the importance of a relatively narrow (1–2 octaves) band of face frequencies located around 8–10 cycles/face (e.g., Costen, Parker, & Craw, 1996; Fiorentini, Maffei & Sandini, 1983; Gold, Bennett, & Sekuler, 1999; but see Hayes, Morrone, & Burr, 1986 who located it to 20 c/face). On the other hand, Tieger and Ganz (1979) concluded that simple channel content was not sufficient to explain their recall task for face recognition and suggested the need for a higher stage of processing. Regarding human face discrimination, it has been suggested that impaired sensitivity to low spatial frequencies is related to poorer face discrimination in a study

* Corresponding author. Address: Department of Psychology, Colgate University, 13 Oak Drive, Hamilton, NY 13346, USA.
E-mail address: bchansen@mail.colgate.edu (B.C. Hansen).
[1] Current address: Department of Optometry and Vision Science, University of Auckland, New Zealand.

0042-6989/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2008.08.016

investigating face discrimination in young and elderly samples (Owsley, Sekuler, & Boldt, 1981). This was subsequently supported by the finding that face discrimination performance is not reduced when faces are significantly blurred or pixelated (White & Li, 2006; but see Goffaux, Hault, Michel, Vuong, & Rossion, 2005). Additionally, Goffaux and Rossion (2006) observed that holistic processing of human faces was dependent on low spatial frequencies.

It seemed somewhat implausible to us that the channel concepts that have been so useful in predicting the detectability of single localized patches of sinusoidal components (Graham, 1980) could, on their own, account for such complex processes as face discrimination or face recognition. Faces are composed of multiple features that are processed in a holistic way (e.g., Schiltz & Rossion, 2006; Sergent, 1984; Tanaka & Farah, 1993; Young, Hellawell, & Hay, 1987). Since the spatial alignment of face features would depend critically on the Fourier phase spectrum (refer to Fig. 1), it is reasonable to assume that such processing would need to rely on this phase-defined content across multiple scales. This process would likely involve the combination of output across individual spatial channels. This would lead one to think in terms of a critical band of phase *alignments* rather than simply a critical octave bandwidth of detectable frequencies for face recognition or discrimination. By phase "alignment", we are referring to the complex Fourier phase relationships of the sinusoidal waveforms of any given image whereby the convergence of arrival phases has been shown to form the edges, lines, or contours of image structure (Morrone & Burr, 1988; Morone & Owens, 1987; see also Hansen & Hess 2007) refer to Fig. 2 for further details. To test this notion, we investigated the spatial range over which face content carried by the phase spectrum needs to be preserved for face discrimination and face recognition at three different peak face frequencies. In addition, we compared human performance with that obtained from template-matching simulations in order to provide a benchmark for interpreting the data.

We used two sets of face stimuli. For face discrimination, we utilized arrays of "face morphs" (Leopold, O'Toole, Vetter & Blanz, 2001), each consisting of samples along a continuum of morphs between a standard "average" face and a given "unique" face. The stimuli therefore, were a series of face images, graded in their strength of identity content. The face morphing technique offers a unique approach for assessing face discrimination thresholds. Consider that face discrimination in its purest sense typically yields categorical-type results (i.e., a sharp step function as opposed to a smooth psychometric function) without offering insight into where along a given identity continuum (between two faces) a given observer is able to reliably discriminate between the two identities. Since morphing allows one to essentially "stretch" identity out along a continuum between two different faces, the use of this technique can more effectively address the question of just how well human observers can discriminate between two faces. The morphing technique therefore yields a reasonably accurate estimate of just how much of a difference in identity content between two faces is needed to successfully discriminate between those faces. We therefore employed morphed images in a partial phase-randomization paradigm whereby each face morph array was subjected to selective phase randomization where variable ranges of the Fourier phases of face frequencies, centered on one of three central face frequencies, were preserved with all other phases randomized. Thus, for a given range of face frequencies, psychometric functions were measured in order to determine the threshold "morph level" (along the morph array) for human observers to successfully discriminate between an average face and different unique faces, thereby providing insight into how much of a change in the content is needed for observers to make successful discriminations. For face recognition, we employed 10 unique face images which were subjected to the same randomization procedure mentioned above. Psychometric functions were measured in order to determine the threshold "bandwidth of phase preservation" for human observers to successfully recognize unique faces.

In the current study, we set out to investigate the utility of phase alignment (i.e., the alignment of different face frequencies) for human face discrimination and recognition in four experiments. Experiment 1 was designed to measure the relative amount of phase alignment needed to successfully discriminate between different faces centered around three different central face frequencies. Human data in Experiment 1 were compared to simulated performance based on cross-correlation in order to evaluate whether such a template-matching model of performance could serve as a useful benchmark for the observed behavioral data. Experiment 2 consisted of a follow-up to Experiment 1 to assess whether the phase-aligned face frequencies present in the smallest phase alignment bandwidths was too weak by itself to drive reliable performance, or whether the reduction in performance observed in Experiment 1 was due to the presence of noise in the form of "misaligned phase components". In the third experiment, we altered the contiguous nature of the phase-preserving filter utilized in Experiment 1 in order to determine if the results in

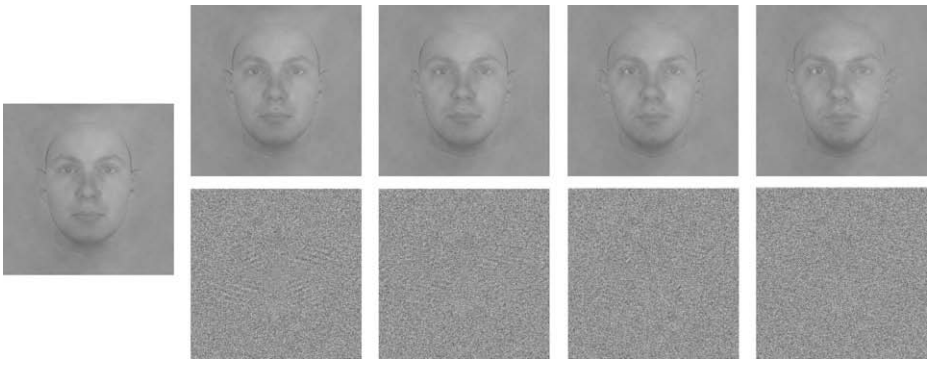

Fig. 1. Illustration highlighting the importance of the Fourier phase spectrum in carrying relevant content for face discrimination. On the far left is a reference face that has had its amplitude spectrum replaced with an isotropic spectrum with an amplitude spectrum fall-off of −1.0. Top row: series of "face morphs" (see text for further details) where the reference face has been gradually "morphed" (from left-to-right) into a different face (top row, far right). All faces in the top row have been assigned the same amplitude spectrum as the reference face. Below each morph face is a normalized "difference phase spectrum" created by subtracting the phase spectrum of the reference face's phase spectrum from the phase spectrum of the corresponding face in the top row. Since the difference spectra have been normalized, values darker than mean gray-level 128 indicate negative differences and values greater than mean gray-level indicate positive differences. Note that, even when the reference face is very similar to a given morph face (e.g., top row, 1st face), the subtle differences in the spatial domain are represented by wide-spread differences between their phase spectra in the Fourier domain.

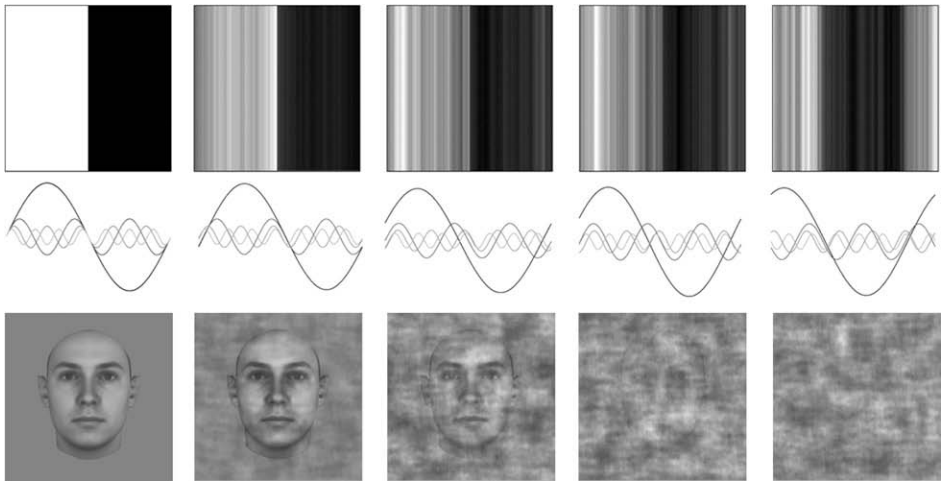

Fig. 2. Top-row: (Top-left) An illustration of an edge, below which is an illustration of a series of sinusoidal waveforms that have arrival phase convergence at the central position which, over a full series of sinusoidal waveforms of increasingly higher spatial frequencies (not shown), will sum up to the edge shown above. The next four illustrations show how the saliency of the edge is corrupted as the alignment of the sinusoidal waveforms is increasingly perturbed. Bottom row: Two-dimensional illustration of how the saliency of the contours in a given 2D image can be disrupted as a function of increasing (left-to-right) phase randomization.

Experiment 1 could be explained by reductions in the global signal-to-noise ratio. And finally, Experiment 4 was constructed to measure the relative amount of phase alignment needed to recognize human faces using similar techniques employed in Experiment 1 (which was designed to measure face discrimination), thereby allowing a direct comparison of the relative amount of phase alignment needed to discriminate between and recognition of different faces. The results from Experiment 4 were also compared to simulated performance based on cross-correlation. Our results provide support for a critical frequency for face recognition (e.g., Costen et al., 1996; Fiorentini, Maffei & Sandini, 1983; Gold et al., 1999; Näsänen, 1999), but not for face discrimination as a function of relative phase alignment. Specifically, our face discrimination results (between an average face and different unique faces) suggest the existence of a fixed *critical band of aligned phases* within which a specific ratio of phase-aligned to phase-misaligned components is required for accurate face discrimination performance, *regardless* of the central frequency at which that range is centered.

2. Experiment 1

In Experiment 1 we were interested in measuring face discrimination thresholds for arrays of face morphs that had been partially phase-randomized. We compared face discrimination thresholds obtained from a condition where the face morph arrays were not subjected to partial phase randomization (i.e., the baseline condition) to conditions where the bandwidth of preserved phase alignment (where all face frequency phases outside the preserved band were randomized) was set at one of five different widths. Each bandwidth was balanced with respect to the number of aligned frequencies it contained. The face morph arrays were taken from those utilized by Leopold et al. (2001), and consisted of morphs between an "average" face and one of four "unique" faces, and can therefore be considered similar to the "within" stimulus category employed by Rotshtein, Henson, Treves, Driver, and Dolan (2005)—the task itself is virtually identical to that used by Beale and Keil (1995), where it was argued that individual faces are perceived categorically. However, here, we were concerned with the amount of Fourier phase alignment needed for observers to reliably discriminate between a common average face and a set of unique faces. The rationale for this approach was twofold: (1) by measuring discrimination thresholds from the average face, the amount of pre-experiment training would be greatly reduced (i.e., instead of training observers on a large set of faces prior to the experiment, they only needed to be familiarized with one face, that being the average face). (2) Since it is likely that human observers might differ in their ability to discriminate between sets of two different morphed identities, we chose to assess the ability to discriminate between unique faces and a fixed face identity, thereby providing a reference that was consistent across all observers.

2.1. Methods

2.1.1. Apparatus

All stimuli were presented with an Intel Pentium IV (3.21 GHz) processor equipped with 1 GB RAM. Stimuli were displayed using a linearized look-up table (generated by calibrating with a UDT S370 Research Optometer) on a 22″ Mitsubishi Diamond Pro 2070^{SB} CRT driven by an ASUS Extreme AX300 Graphics card with 8-bit grayscale resolution. Maximum luminance was 100 cd/m^2, the frame rate was 120 Hz, and the resolution was 1600 × 1200 pixels. Single pixels subtended .013° visual angle (i.e., 0.78 arc min.) as viewed from 1.0 m.

2.1.2. Participants

Five psychophysical observers participated in the current experiment, two experienced psychophysical observers (one naïve to the purpose of the study) and three relatively experienced observers (undergraduates that had some experience participating in psychophysical experiments and were naïve to the purpose of the experiment). All participants had normal (or corrected-to-normal) vision. The ages of the participants ranged from 20 and 31 years. Research Ethics Board-approved informed consent was obtained.

2.1.3. Stimulus construction

The face stimuli utilized in the current experiment consisted of four "face morph arrays" sampled from the face space described in Leopold et al. (2001), and consisted of the same four face arrays shown in that study. However, for the current study, we were only interested in the morph series between the "average" face and each of the four "unique" faces (see Leopold et al., 2001 for further details). The full color face images themselves were provided to us by David Leopold and were generated from the face database of the

Max Planck Institute for Biological Cybernetics (Troje & Bülthoff, 1996). The face morph vectors were generated by a face morphing algorithm where point-by-point locations between face pairs were matched using 3D structure and reference maps (Blanz & Vetter, 1999) which gradually morphed the average face into one of the four unique faces in linear steps from 0% unique (i.e., the average face) to 100% unique. In order to bring the four sets of face morphs inline with the current experiment, all face images were padded in order to make the total image dimensions 400 × 400 pixels (each face was centered, with an average ear-to-ear width of 255 pixels, *SD* = 9 pixels) and were converted to grayscale using the standard NTSC (National Television Standards Committee) formula (i.e., luminosity = 0.299 * $R(x)$ + 0.587 * $G(x)$ + 0.114 * $B(x)$). Next, each grayscale face was normalized to the range [0,1] and locally (local with respect to each face area) assigned the same mean luminance (normalized grayscale 0.5) and rms contrast (normalized 0.09), with all pixels falling outside of the face area assigned the same normalized mean luminance (i.e., 0.5).

In order to test the amount of "phase alignment" needed for human observers to make face discrimination judgments, systematic selective randomization of the face images' phase spectra was carried out in the Fourier domain. Using MATLAB (version 7.0.4) and accompanying Image Processing and Signal Processing Toolboxes (versions 5.0.3 and 6.3, respectively), each image was, in turn, subjected to a discrete Fourier transform which yielded an amplitude spectrum, $I_{AMP}(f,\theta)$ and phase spectrum, $\Phi(f,\theta)$, where f and θ represent given spatial frequency and orientation locations, respectively, in polar coordinates.

In order to assure that the stimulus images differed only with respect to their phase spectra, an isotropic amplitude spectrum was generated with the amplitude fall-off typical of natural scenes i.e., $1/f^{\alpha}$, where $\alpha = 1.0$ (Billock, 2000, Burton & Moorhead, 1987; Field, 1987; Hansen & Essock, 2005; Ruderman & Bialek, 1994; Tolhurst, Tadmor, & Tang Chao, 1992; van der Schaaf & van Hateren, 1996). The isotropic spectrum, $ISO_{AMP}(f,\theta)$, was generated by constructing an empty matrix of the same dimensions as the stimulus imagery and assigning each spatial frequency coordinate (in *polar coordinates*) a value from the following function:

$$ISO_{AMP}(f_i,\theta_j) = \frac{1}{f_i^{\alpha}} \quad (1)$$

Thus, for each f coordinate along the radius axes, the same value is assigned to each orientation. Using this isotropic spectrum ensured that the only difference between the images would be in their phase spectra, as well as helping to reduce any "edge effects" that may have arisen during the initial Fourier transform.

Next, the systematic selective randomization algorithm was implemented by preserving the phases within a range of spatial frequencies for all orientations. The phase spectra were filtered in polar coordinates, with an "ideal filter" which was defined as follows:

$$PP_{FILT}(f_i,\theta_j) = \begin{Bmatrix} \Phi(f_i,\theta_j) & f_L \le f_i \le f_H \\ \text{rand}([-\pi,\pi]) & \text{elsewhere} \end{Bmatrix} \quad (2)$$

where f_L and f_H are the lower and upper spatial frequency bounds of the phase preserving "ideal filter" (PP_{FILT}). This filter preserves the phase angles of a given phase spectrum for the frequencies ranging from f_L to f_H (i.e., preserves the original phase angles falling within the pass-band of the filter) and assigns a random value ranging from $-\pi$ to π to the coordinates falling outside of this range (refer to Hansen & Hess, 2007 for further details). Thus, PP_{FILT} preserves the "aligned" relationship shown in the left-most panel of Fig. 2 for frequencies falling within its pass-band and perturbs that relationship (e.g., right-most panel of Fig. 2) for frequencies falling outside its pass-band. Note that the odd symmetry of the phase spectrum was maintained (not represented in Eq. (2)) i.e., for θ angles in the $[\pi,2\pi]$ half of polar space, $PP_{FILT}(f,\theta) = PP_{FILT}(f,\theta-\pi)^{*}(-1)$. For all of the experiments in the current study, PP_{FILT} was centered on one of three different spatial frequencies, calculated with respect to cycles per face. In order to target the appropriate face frequencies in the Fourier domain, the desired face frequency needed to be converted to cycles/picture. Specifically, the cycles/face-to-cycles/picture calculation involved multiplying the desired cycle/face frequency by the ratio of the total image width to the average ear-to-ear face width in pixels. For the current study, the central face frequencies of PP_{FILT} were 3, 9, and 27 cpf. For each of the three central face frequencies, the bandwidth of PP_{FILT} was fixed at five different *octave* widths (or "levels", where, for level one, 3 cpf = 2 – octaves; 9 cpf = 0.6 – octave; and 27 cpf = 0.3 – octave) for each of the central face frequencies: 3 cpf [2.0, 3.0, 3.5, 4.0, 5.0], 9 cpf [0.6, 1.0, 1.5, 2.0, 3.0], and 27 cpf [0.3, 0.6, 1.0, 1.5, 2.0], resulting in five "levels" of phase alignment for each of the central face frequencies. The rationale for having different octave bandwidths for each central frequency was that since spatial frequency is assessed with respect to "cycles per face" (a fixed measure which does not change with viewing distance), an octave of face frequencies centered on 3 cpf will only contain face frequencies from the 2- to 4-cpf range, whereas an octave centered on 9 cpf will contain face frequencies from the 6- to 12-cpf range (i.e., more frequencies in the latter case means more alignment). Once a given face image's phase spectrum had been filtered, it was assigned a copy of the isotropic amplitude spectrum described above, and subjected to an inverse discrete Fourier transform (refer to Fig. 3 for examples).

It has been argued that when comparing face identification performance across different central frequencies, it is important to correct for any differences in the amount of "information" available to observers. Such corrections have been achieved with the use of an ideal observer model (e.g., Gold et al., 1999). Here, the bandwidths for the three different central face frequencies were not identical in order to keep the number of phase-aligned frequencies comparable across each of the three different central face frequency conditions. This did not directly ensure that there was approximately the same amount of "information" available within each bandwidth level, however here we were concerned with the

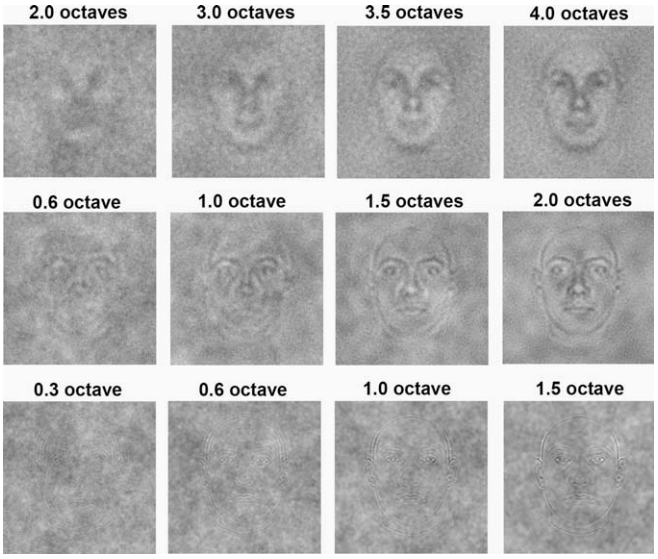

Fig. 3. Example stimuli from Experiment 1. Each row corresponds to each of the central frequencies upon which PP_{FILT} was centered: 3, 9, and 27 cpf, respectively. Each column corresponds to the bandwidth of PP_{FILT} (i.e., the range of non-randomized phase angles as a function of face frequency) in octaves for each of the central face frequencies.

relative amount of phase-aligned frequencies needed to discriminate or recognize human faces based on phase-aligned content and reserve the issue of assessing available information for stimuli such as those described above for another study. In this context, a bench mark to which human performance could be compared was provided by conducting multiple template-matching simulations (described below). Template matching was chosen as it not only provided a benchmark for psychophysical performance, but allowed for a direct test of whether this was the strategy employed by the human observers.

2.1.4. Psychophysical procedure

The general psychophysical task in the current experiment employed a 2AFC method of constant stimuli paradigm. Participants viewed the display monitor at a distance of 1 m; face stimuli subtended $3.1^\circ \times 4.18^\circ$ visual angle. For any given trial, observers were presented with a fixation dot (0.82° visual angle) placed at the center of the display monitor (500 ms), followed by stimulus interval 1 (200 ms), followed by a $1/f$ noise mask (500 ms), followed by the fixation dot (500 ms), followed by stimulus interval 2 (200 ms), followed by another $1/f$ noise mask (500 ms), followed by an empty display (set to mean luminance) where the observers were required to make a response via key-press (the duration of the response interval was unlimited). Stimulus interval duration was chosen to match that used by Leopold et al. (2001). Feedback was not provided, and all stimuli were viewed binocularly. The task of the observers was to indicate which of the two stimulus intervals *did not* contain the average face (stimulus interval order was random). Note that a different random seed was used for the phase randomization employed to generate the stimuli on each trial to ensure that observers engaged in face-based discrimination rather than an "image-based" (i.e., "noise-based") discrimination.

Prior to the start of the current experiment (1 day prior), all observers were asked to familiarize themselves with the average face (non-phase filtered) and were allowed to re-familiarize themselves any time between experimental sessions (though none needed to do this). All experimental sessions were grouped by unique face. Within each session, observers were required to discriminate between the average face and 10 morphed faces along the face morph array between the average face and the unique face (one morphed face presented per trial, selected randomly). For the current study, we chose face morphs ranging from 5% unique to 50% unique (in 5% face morph steps), and all sessions were repeated twice. All observers were allowed practice sessions to familiarize themselves with the task prior to engaging in the experimental sessions. Threshold estimates for successful face discrimination were assessed using "psignifit" (Wichmann & Hill, 2001a, 2001b), using the Weibull fit option, and were calculated for performance at each PP_{FILT} octave width and central face frequency. Threshold estimates were taken for the data averaged across the four unique faces (preliminary data suggested very similar performance for each unique face) for each observer and then averaged across all observers.

2.1.5. Template-matching benchmark

In Experiment 1, for any given trial, the task was to identify which of two sequentially presented faces (both partially phase-randomized) was not the average face. Thus, performance for such a task could be explained by simple "template matching" (i.e., spatial cross-correlation). Such a template-matching strategy could use the broadband (i.e., all spatial frequencies and orientations preserved) average face as a reference "template" that is cross-correlated with each stimulus interval in each trial. A correct response could be given according to which stimulus interval yielded the lowest cross-correlation. Similar approaches have been shown to be effective for the identification of complex objects and faces (e.g., Braje et al., 1995; Gold et al., 1999; Tjan, Braje, Legge, & Kersten, 1995). To account for this possibility, template matcher performance was assessed by simulating the experiment described above, where, for each trial, either the non-filtered average face (or a band-pass amplitude-filtered version) was cross-correlated with each stimulus interval. In order to maximize the cross-correlation signal, only face areas were cross-correlated, not the total image area. The template matcher responded by selecting the lowest cross-correlation on each trial as the interval that did not contain the average face. Estimates of the template matcher's thresholds were made using the same procedures described in the psychophysical procedure section of the current experiment. The experiment was simulated five times in order to have one template matcher for each human subject that participated in the current experiment.

2.2. Results

Prior to participating in Experiment 1, all participants were measured for their ability to discriminate non-phase-spectrum-filtered morphed faces from the average face, utilizing the same psychophysical paradigm described above. The average face discrimination threshold (with respect to the % of unique face content needed to successfully discriminate a given morphed face from the average face) was 19.55% unique face, $SE = 2.37\%$. All subsequent analyses will be made with respect to this discrimination threshold obtained from the non-phase-filtered control. The results from Experiment 1 are shown in Fig. 4. Each of the inset graphs of that figure show the psychometric functions for each of the five different bandwidths of PP_{FILT} centered at each of the three different central face frequencies. For each central face frequency, there is a clear improvement in face discrimination threshold as the bandwidth of PP_{FILT} increases. This effect stabilizes around a morph level of 20% unique face. This observation is made explicit in Fig. 5a which shows a plot of the averaged morph-thresholds for each of the three different central face frequencies across each of the five PP_{FILT} bandwidths. Data were analyzed with a 3 (central face frequency) $\times$ 5 (PP_{FILT} bandwidth) two-way repeated-measures Analysis of Variance (ANOVA) using a reasonably conservative correction (i.e., Huynh–Feldt epsilon) to adjust the degrees of freedom (Cohen, 2001). The main effect of central frequency was not significant ($F_{1,4} = 0.194$, $p > .05$) indicating that the thresholds across the three different central face frequencies were similar. The main effect of PP_{FILT} bandwidth was significant ($F_{2,6} = 11.87$, $p < .01$), indicating thresholds depended on PP_{FILT} bandwidth. The interaction was not significant ($F_{2,6} = 0.43$, $p > .05$) indicating a similar pattern of threshold data across PP_{FILT} bandwidth for the three central face frequencies. Post-hoc paired t-tests (assuming un-equal variances) showed that the differences between thresholds for each of the last three filter bandwidth levels across each of the three central frequencies were not significant ($p > .05$).

For performance to reach the baseline level (i.e., 20% unique face), the participants needed different octave bandwidths which were broadest for the 3 cpf condition and narrowest for the 27 cpf condition, suggesting that human observers require less phase alignment at higher face frequencies than lower face frequencies, with the middle frequencies being intermediate. However, since we chose bandwidths that contained an approximately equivalent number of aligned frequencies for each bandwidth "level", the most informative way to show the data is with respect to the number of aligned frequencies within each bandwidth level. It should be noted that when PP_{FILT} was centered at 27 cpf, the spatial frequencies at and above 27 cpf were beyond the spatial frequency resolution limit of the visual system (e.g., Campbell & Green, 1965). Accordingly, it was necessary to consider only the spatial frequencies within the lower bound (i.e., at and

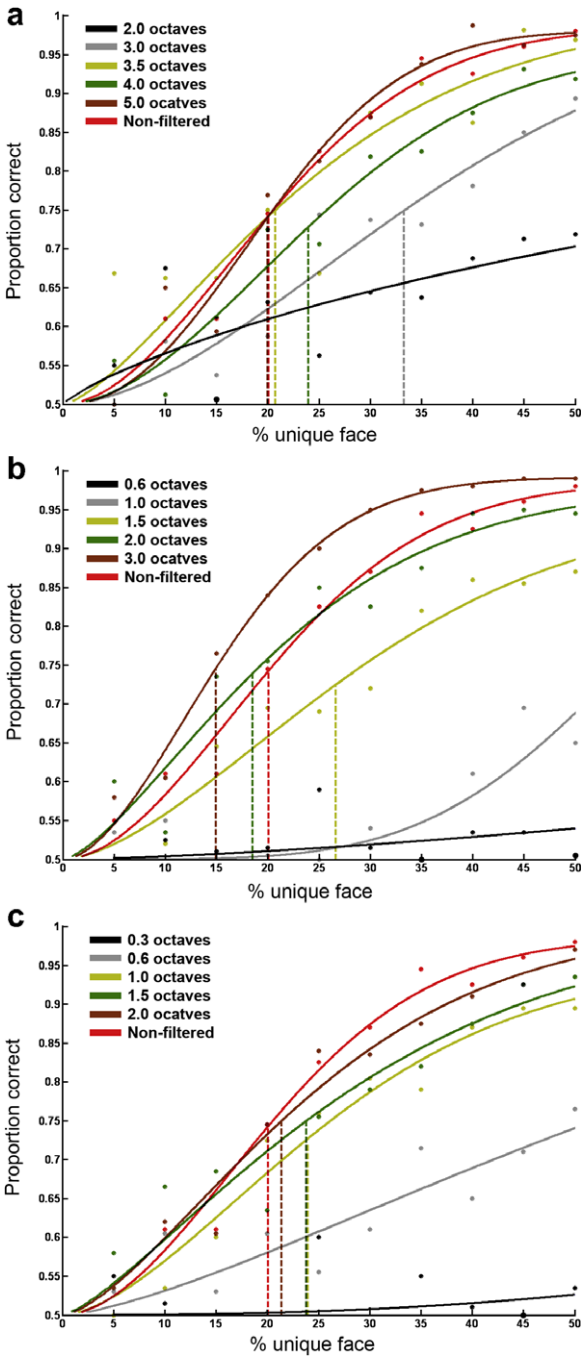

Fig. 4. Averaged data from Experiment 1. Each graph shows data, along with their fitted psychometric functions, for each of the three different central face frequencies at which PP_{FILT} was centered: (a) 3, (b) 9, and (c) 27 cpf. On the ordinate of each graph is averaged proportion correct (averaged across all observers and four face morph arrays), and on the abscissa of each graph is the % of the unique face for the four different face morph arrays. The light-red data plotted in all three graphs are from the non-phase randomized face stimuli condition, shown for comparison.

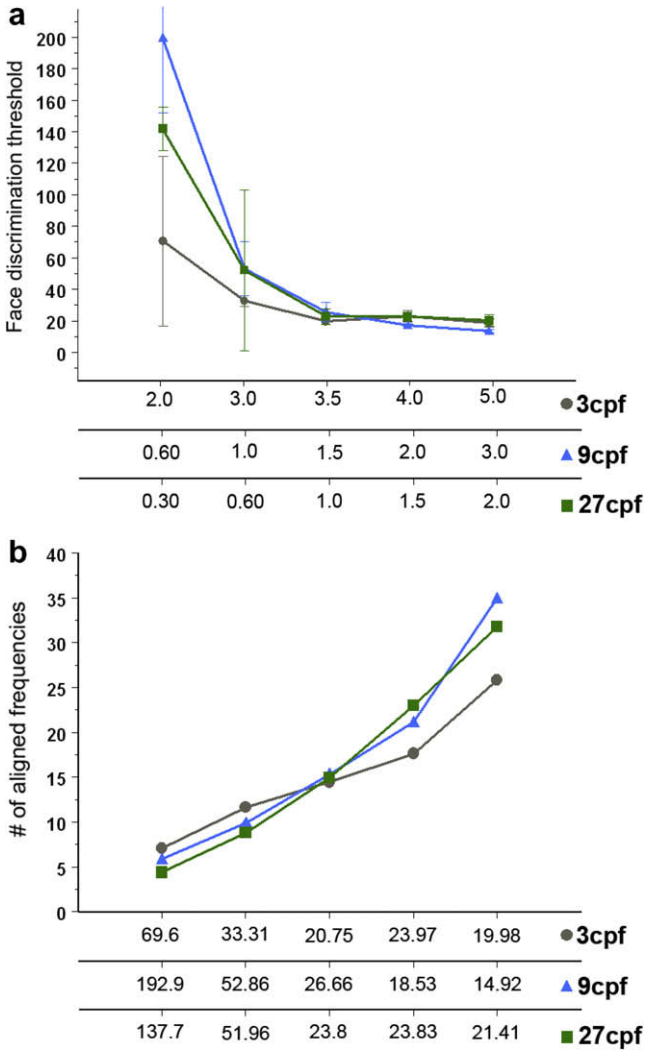

Fig. 5. Data re-plotted from Experiment 1. (a) On the ordinate is averaged face discrimination threshold, and on the abscissa are the five different PP_{FILT} bandwidths for each central face frequency. Error bars are ±1*SEM* (calculated between observers). Refer to the text for further details. (b) On the ordinate is number of phase-aligned face frequencies within the pass-band of each of PP_{FILT}. On the abscissa are the averaged face discrimination thresholds for each of the central face frequencies of PP_{FILT} as a function of bandwidth (as plotted in Fig. 5a). Note that some of the threshold estimates are outside of the range of the morphed stimuli and should not be considered to be accurate, rather they reflect the general difficulty participants had making discriminations on those trials.

below 27 cpf) of PP_{FILT} when it was centered on 27 cpf, that is, only the alignments below the central frequency were summed. After accounting for that caveat, the number of aligned frequencies within each bandwidth of PP_{FILT} were plotted as a function of face discrimination threshold. Fig. 5b clearly shows that for observers to reach a face discrimination threshold comparable to that obtained with the non-phase-spectrum-filtered stimuli (i.e., ~20% unique face), the same number of aligned frequencies (~16–18 phase-aligned frequencies) is needed regardless of the central frequency of PP_{FILT}. The data argue that the amount of phase alignment needed to successfully discriminate between the content of different faces (i.e., when compared to non-phase-randomized control face discrimination thresholds) is approximately constant (i.e., independent of face frequency). While outside the focus of our analysis, it is worth noting that there does appear to be a performance bias in favor of the lowest central face frequency (3 cpf) condition for the narrowest PP_{FILT} bandwidth. That is, the averaged thresholds do appear to be somewhat lower than the other two central face frequencies which is somewhat consistent with previous face discrimination literature (Goffaux & Rossion, 2006; Owsley et al., 1981; White & Li, 2006).

The averaged threshold data from the template matchers are plotted in Fig. 6 along with the threshold data of the human subjects (re-plotted from Fig. 5a) for comparison. As shown in Fig. 6a, the template matchers show the same trend in discrimination threshold as the human observers, with thresholds becoming increasingly smaller as PP_{FILT} increased in bandwidth. However, in

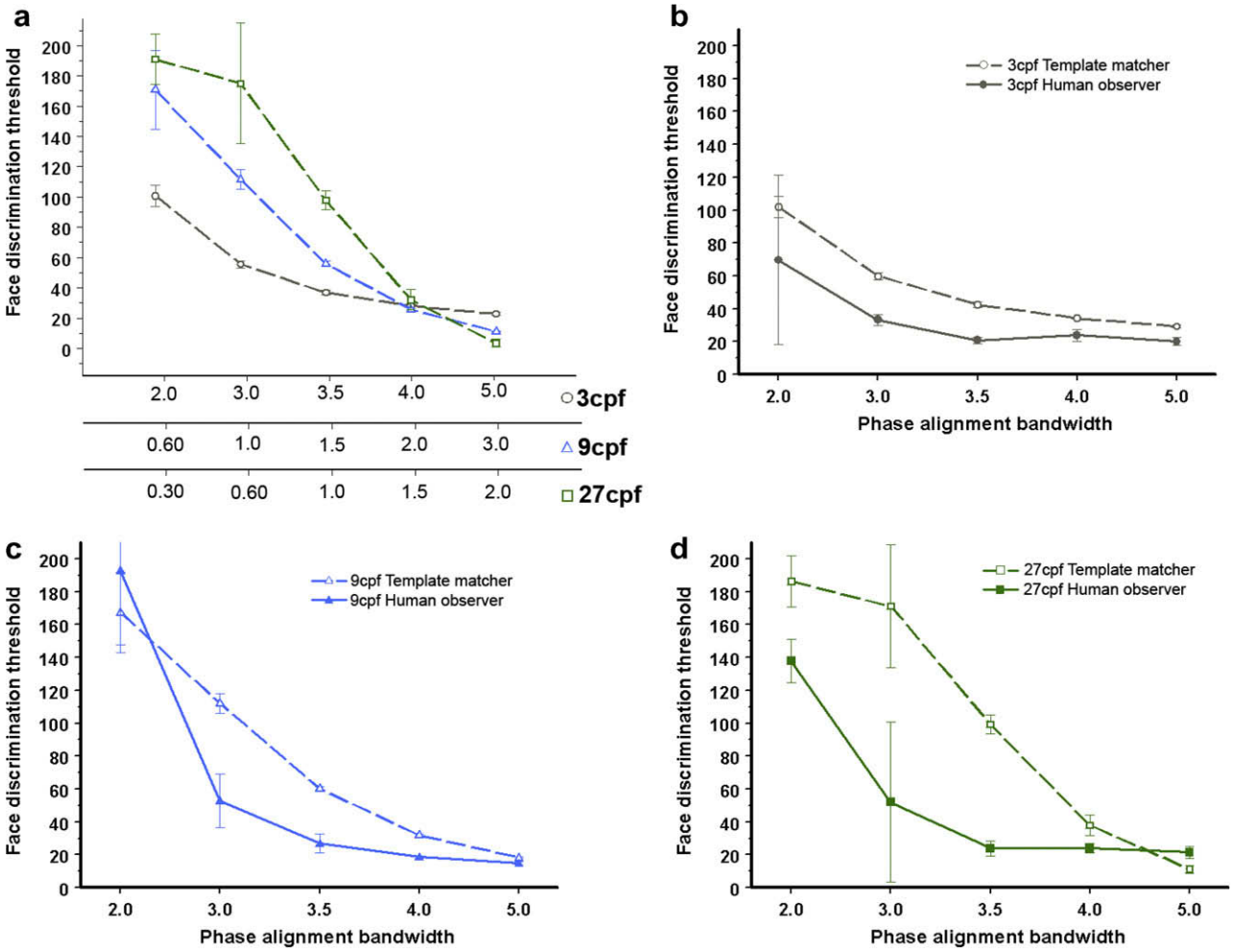

Fig. 6. Data re-plotted from Experiment 1. (a) Averaged ideal observer performance, figure axes are identical to those in Fig. 5a. (b) Human observer data from Experiment 1 re-plotted along with the ideal observer data for the 3 cpf central face frequency condition. (c) Human observer data from Experiment 1 re-plotted along with the ideal observer data for the 9-cpf central face frequency condition. (d) Human observer data from Experiment 1 re-plotted along with the ideal observer data for the 27 cpf central face frequency condition. Refer to text for further details.

all three central frequency conditions (i.e., Fig. 6b–d), the human observers out-performed the template matchers up to the broadest PP_{FILT} bandwidth, indicating that such a strategy is not employed in human face discrimination based on phase alignment. Thus, in the range where the human observers' thresholds began to approach those obtained in the control condition, they were able to make face discriminations more effectively than simple template matching would predict when the template made use of all available face frequencies. In order to determine whether the thresholds for the human observers and the template matchers differed, a 2 (human vs ideal) × 3 (central face frequency) × 5 (PP_{FILT} bandwidth) three-way ANOVA was conducted. There was a significant main effect between human and template matcher thresholds ($F_{1,7} = 8.3$, $p < .05$), indicating that the two groups' thresholds were not identical.

We also conducted a template matcher simulation where the "face template" consisted of a band-pass amplitude-filtered version of the average face, with the peak frequency of the band-pass filter centered on one of the three nominal face frequencies employed in the current experiment (the method for band-pass amplitude filtering used here is described in the method section of Experiment 2—in short, no phase randomization was applied to the template), data not shown. The simulation was identical to that described above, except with different band-pass filtered templates. When the spatial frequency bandwidth of the template was set to 1 octave, the template matcher performed the task perfectly for all three central face frequency conditions. It was not until the bandwidth was increased beyond 2–3 octaves that the template matcher began making errors in all three central face frequency conditions, suggesting that if humans do indeed employ a template-matching strategy for discriminating faces based on phase alignment, the template would likely consist of a broad range of face frequencies (but not all face frequencies—see Fig. 6) as it is at those bandwidths where the template matcher began to perform similarly to the observed human performance in the current experiment.

3. Experiment 2

The results of Experiment 1 argue against a critical bandwidth for face discrimination with respect to different central face frequencies. A major factor that influenced a reduction in performance would have been the number of the phase-aligned frequencies, rather than the central face frequency of PP_{FILT}. Here, as a follow-up to Experiment 1, we sought to evaluate whether the phase-aligned frequencies present in the smallest bandwidth was too weak by itself to drive reliable performance, or whether

the reduction in performance observed in Experiment 1 was due to the presence of noise in the form of "misaligned phase components". In order to differentiate between those two possibilities, we repeated Experiment 1 using band-pass filtered face stimuli. Accordingly the *amplitude spectra* of the face morph arrays were filtered with a log-Gaussian filter fixed at three different bandwidths, centered at two different central face frequencies (note that the stimuli in the current experiment were *not* filtered with PP_{FILT}).

3.1. Methods

3.1.1. Apparatus

Same as in Experiment 1.

3.1.2. Participants

Two experienced psychophysical observers participated in the current experiment. Both participants had normal (or corrected-to-normal) vision. The ages of the participants were 29 and 31 years. Research Ethics Board-approved informed consent was obtained.

3.1.3. Stimulus construction

Stimuli for the current experiment were constructed from the same face image set used in Experiment 1. However, here we band-pass filtered the *amplitude spectra* of those images using a smooth spatial frequency filter in the Fourier domain. Ideally, one would use a Gaussian filter (Gaussian along the frequency axes in polar coordinates) with a given bandwidth centered at a given spatial frequency. Unfortunately, the Gaussian function tends to overlap (i.e., is cropped) at the DC component of the amplitude spectrum when centered on lower spatial frequencies. A reasonable solution to this issue is to use a log-Gaussian function, which will always approach zero near the DC component. In the Fourier domain, the log-Gaussian filter can be expressed as:

$$L_{GAUS}(f,\theta) = e - \left[\frac{\log\left(R(f_i,\theta_j)/F_{peak}\right)^2}{2\log\left(\sigma_1/F_{peak}\right)^2}\right] \quad (3)$$

where f_i and θ_j represent any given position in polar coordinates, R represents a given radius vector (i.e., the spatial frequency dimension) taken from $I_{AMP}(f,\theta)$, F_{peak} is the central spatial frequency of the log-Gaussian function, σ_1 is the spatial frequency bandwidth of the log-Gaussian function.

Using the same discrete Fourier functions mentioned in Section 2.1, each face image was, in turn, Fourier transformed and its amplitude spectrum filtered with L_{GAUS} centered at either 6 or 18 cpf. It was not practical to center L_{GAUS} on extreme lower and upper face frequencies (i.e., 3 and 27 cpf) due to sampling limitations in the Fourier domain at those frequencies. Since the aim of the current experiment was simply to evaluate whether phase-aligned content present in the smallest bandwidth was too weak by itself to drive reliable performance, specific central frequency was not important. For each central face frequency tested here, three different bandwidths of L_{GAUS} were implemented. When L_{GAUS} was centered on 6 cpf, the bandwidths included 0.38, 0.53, and 0.80 octaves (full-width at half-height), and when it was centered on 18 cpf, the bandwidths included 0.27, 0.42, and 0.81 octaves (full-width at half-height). Unfortunately, due to L_{GAUS} becoming extremely peaked (and hence less smooth, i.e., resembling a delta function) for very narrow bandwidths, we were limited as to how narrow we could make L_{GAUS}. Once each amplitude spectrum had been filtered, it was inverse discrete Fourier transformed with its respective phase spectrum back to the spatial domain. In the spatial domain, each image was assigned the same mean luminance and rms. contrast used in Experiment 1. Thus, for each central face frequency, there were three sets of face stimuli (see Fig. 7 for examples), resulting in six different experimental sessions. The stimuli generated here differed from Experiment 1 in that they did not contain any of the noise created by randomizing phases outside of pass-band of the filter—they did possess similar relative phase alignments, but filtered out the contrast energy for alignments outside the filter bandwidth. The experimental paradigm itself was identical to that used in Experiment 1 (that is, for each given trial, observers were required to discriminate between the average face and one of the unique face morphs, where both the average and unique morph were identically filtered).

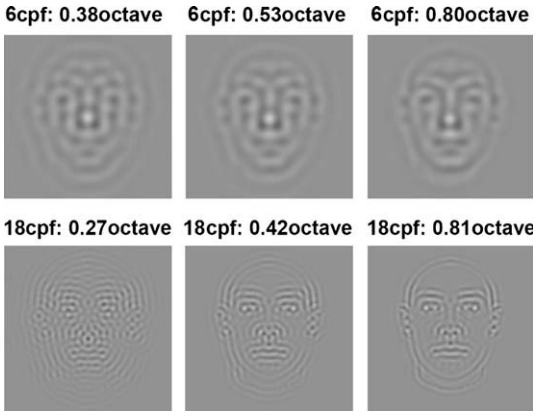

Fig. 7. Examples of some of the stimuli used in Experiment 2. Each row corresponds to the two different central face frequencies of L_{GAUS}, and each column is the bandwidth in octaves (half-width at half-height) of L_{GAUS}.

3.2. Results

The results for both observers are shown in Fig. 8. Both observers were able to discriminate between morphed faces and the average face with discrimination thresholds near those obtained with the non-filtered face image set with very limited L_{GAUS} bandwidths indicating that, for a fixed suprathreshold rms contrast of 0.09, human observers can discriminate between faces possessing extremely narrow ranges of amplitude coefficients which is quite different from what was reported in Experiment 1. The results from both participants were not in total agreement however. There was a reduction in performance by BH at .38 octave (6 cpf) and both BH and RF at .27 octave (18 cpf) as compared to the other two octave widths for both conditions. However,as shown in Fig. 7, faces filtered at the narrowest filter widths hardly resemble faces, and thus performance may have been dictated by a mechanism other than face discrimination such as pattern discrimination. Alternatively the reduction in performance may be due to putative "expert" face systems being unable to effectively process these specific stimuli. The broadest bandwidths used in the current experiment are likely to have engaged the normal face discrimination mechanisms, even though the bandwidths used were mostly narrower than any of the bandwidths used in Experiment 1. The current results suggest that the critical factor impairing performance in Experiment 1 was not simply the reduction in the number of phase-aligned frequencies, but also the presence of noise in the form of misaligned phase components. Thus it appears that one factor limiting performance in Experiment 1 was the presence of misaligned face frequencies in a given targeted channel.

4. Experiment 3

The performance reductions found in Experiments 1 and 2 may be due to (1) a global signal-to-noise reduction that occurs when

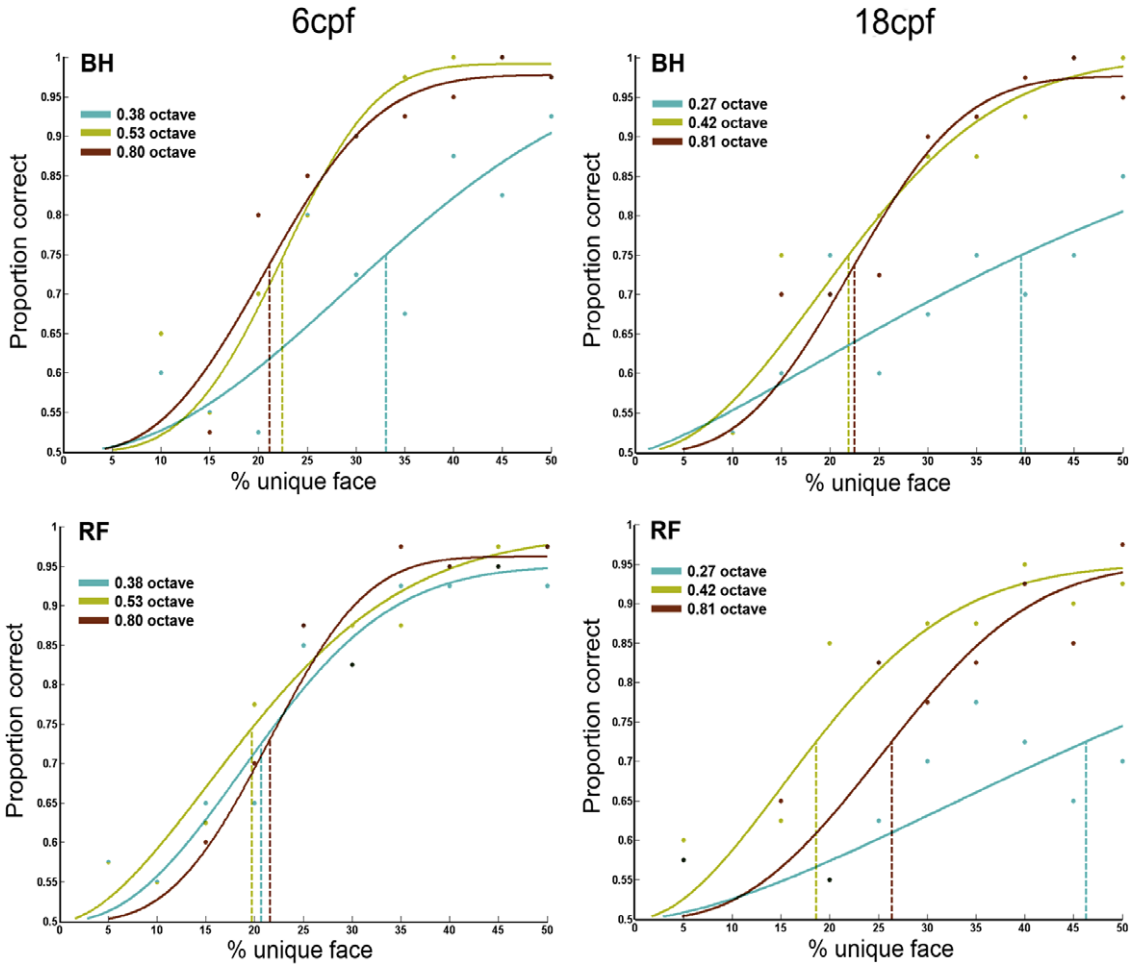

Fig. 8. Individual observer data from Experiment 2, averaged across the four face morph arrays. Each row corresponds to the two observers (B.H. and R.F.) and each column corresponds to the two different central face frequencies upon which L_{GAUS} was centered. On the ordinate of each graph is averaged proportion correct, and on the abscissa of each graph is the % of the unique face for the four different face morph arrays.

bandwidths are reduced, or (2) the reduction in signal-to-noise ratio for a given contiguous band of frequencies. Expanding on the second possibility, it would be expected that distributing a fixed number of phase-aligned frequencies over the frequency domain would impair performance whilst keep the signal-to-noise ratio in the global image constant. In order to investigate these possibilities further, we constructed a new set of selective phase-scrambled face images, where the total number of preserved frequency phases was constant (held at 18 frequencies) while the spacing between those frequencies was varied.

4.1. Apparatus

Same as in Experiments 1 and 2

4.2. Participants

The same two psychophysical observers that participated in Experiment 2 also participated in Experiment 3.

4.3. Stimulus construction

In order to hold the *total* number of phase-aligned frequencies constant, while varying the frequency-to-frequency alignments across a range of spatial frequencies (i.e., disrupting the contiguousness of the filter used in Experiment 1), a "comb-filter" was utilized. The method for selective phase randomizing in a comb-filter manner can be expressed in polar coordinates as:

$$\mathrm{PP}_{\mathrm{COMB}}(f_i,\theta_j)=\begin{Bmatrix}\Phi(f_i,\theta_j) & f_{i,i+1};\quad i=i+\mathrm{GAP}\\ \mathrm{rand}([-\pi,\pi]) & \text{elsewhere}\end{Bmatrix} \tag{4}$$

where GAP specifies the distance between each range of preserved phase angles. Specifically, the "teeth", denoted as $f_{i,i+1}$, of the comb-filter, $\mathrm{PP}_{\mathrm{COMB}}$, consisted of two sequential frequencies in the phase spectrum, and the "gap" between the teeth was variable. $\mathrm{PP}_{\mathrm{COMB}}$ operates in an identical fashion as $\mathrm{PP}_{\mathrm{FILT}}$, except gaps are introduced into the phase-preserved pass-band while increasing the total "gapped bandwidth" of the filter. Note that the total number of phase-preserved frequencies remains constant with this filter even though the total "gapped bandwidth" increases. Examples of face images filtered with PP_{COMB}, as well as an illustration of the filter, are shown in Fig. 9. In that figure, it is clear that with a comb-filter gap of just four frequencies, most of the discernable structure is destroyed, thus arguing against the results of Experiments 1 and 2 being explained by a global image signal-to-noise ratio reduction. In order to verify this observation objectively, we carried out the same psychophysical paradigm described in Experiments 1 and 2, but with a face image set that had been filtered by $\mathrm{PP}_{\mathrm{COMB}}$, where GAP was set at either 0, 2, or 4 frequencies. The data are shown in Fig. 10, and clearly show that when GAP = 4, the task was rendered virtually impossible. These results lend support to the idea that the effects observed in Experiments 1 and 2 are likely due to

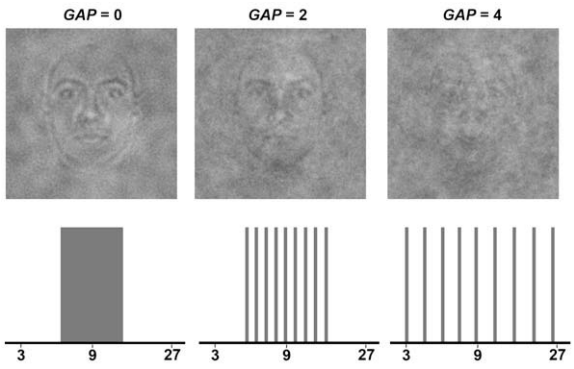

Fig. 9. Stimulus examples and illustrations of the filter employed in Experiment 3. Along the top row are examples of some of the stimuli generated by filtering with PP_{COMB}, with the GAP parameter set at three different values. Along the bottom are 1D illustrations of PP_{COMB} for the images directly above. The illustrations are not to scale, and the "magnitude" of the filter illustrations are completely arbitrary. Note that as GAP increases, the summed total *filter* area is constant, yet the "saliency" of the faces is dramatically reduced.

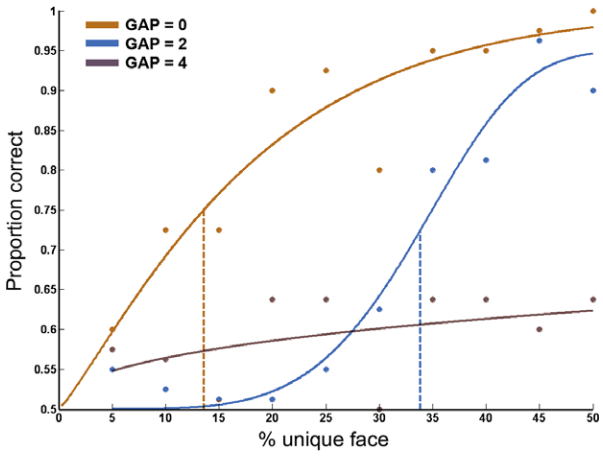

Fig. 10. Averaged data from Experiment 3 (averaged across the two observers and all face morph arrays). On the ordinate is averaged proportion correct, and on the abscissa is the % of the unique face for the four different face morph arrays.

the reduction in the signal/noise ratio for a given *contiguous* band of frequencies.

5. Experiment 4

The results from Experiment 1 demonstrated that the ability of human observers to successfully discriminate between faces critically depends on the number of aligned frequencies and that this dependency operates as a function of the signal-to-noise ratio of phase alignments within a *contiguous* range of face frequencies, regardless of its location in the face frequency spectrum. This finding contradicts a large proportion of the literature devoted to face processing which champions the idea of a critical face frequency (i.e., ~8 to 10 cpf) for face recognition. However, previous studies investigating the spatial processes underlying face perception have primarily focused on face recognition tasks with face stimuli consisting of low-, high-, or band-passed face images where the amplitude spectra have been filtered according to octave bandwidths. The methodology employed in the current study differs from that used in previous studies in that (1) our tasks involved the ability of humans to *discriminate* between different faces on a trial-by-trial basis and (2) we were interested in the role of phase alignment (specifically, the number of phase-aligned frequencies) in face discrimination. The latter has been investigated by Näsänen (1999) in a face recognition paradigm. In that study it was shown that the ability of humans to recognize faces was dramatically reduced when the phases of a band of frequencies were randomized near a central face frequency of 10 cpf. While there exist a number of methodological differences between that study and the experiments described here, the most likely explanation for the differences lies in the fact that the observers in Näsänen's (1999) study and the current study were engaged in fundamentally different strategies, specifically recognition as opposed to discrimination. We therefore set out to resolve this issue in the current experiment by applying the same phase-randomizing methodology utilized in Experiment 1 to a face recognition paradigm.

5.1. Methods

5.1.1. Apparatus

All stimuli were presented with an Intel Pentium IV (2.4 GHz) processor equipped with 1 GB RAM. Stimuli were displayed using a linearized look-up table on a 22″ NuVision 21 MX-SL CRT driven by a VSG2/5 graphics card (Cambridge Research Systems) with 15-bit grayscale resolution. Maximum luminance was 80 cd/m^{-2}, the frame rate was 120 Hz, and the resolution was 1024 × 768 pixels. Single pixels subtended 0.053° visual angle (i.e., 3.18 arc min.) as viewed from 27 cm.

5.1.2. Participants

Four psychophysical observers participated in Experiment 4. Two experienced psychophysical observers (one naïve to the purpose of the study) and two relatively experienced observers (undergraduates that had some experience participating in psychophysical experiments). All participants had normal (or corrected-to-normal) vision. The ages of all participants ranged between 20 and 32. Research Ethics Board-approved informed consent was obtained.

5.1.3. Stimulus construction

The face stimuli were sampled from the face database of the Max Planck Institute for Biological Cybernetics (Troje & Bülthoff, 1996), available at http://faces.kyb.tuebingen.mpg.de/. Five female and five male faces were pseudo-randomly sampled (sampled such that the horizontal/vertical dimensions and luminance contrast were approximately equivalent). All face images were padded in order to make the total image dimensions 256 × 256 pixels (each face was centered, with an average ear-to-ear width of 170 pixels, *SD* = 7 pixels) and were converted to grayscale using the same formula employed in Experiment 1. Next, each grayscale face was normalized to the range [0, 1] and locally (local with respect to each face area) assigned the same mean luminance (normalized grayscale 0.5) and rms contrast (normalized 0.09), with all pixels falling outside of the face area assigned the same normalized mean luminance (i.e., 0.5). Each face was assigned an arbitrary name that the human observers were to use in identifying each face (examples of the 10 faces are shown in Fig. 11). The same partial phase randomization procedure taken in Experiment 1 was also applied here (Eq. (2)). For the current study, the central face frequencies of PP_{FILT} were 3, 9, and 23 cpf. For each of the three central face frequencies, the bandwidth of PP_{FILT} was fixed at 19 different PP_{FILT} bandwidths ranging from 8 to 44 in steps of 2 (full-width). After filtering, each face image was assigned an isotropic amplitude spectrum (Eq. (1)) and subsequently inverse Fourier transformed. In the spatial domain, all stimuli were set to the same mean luminance and rms contrast as the stimuli used in Experiment 1.

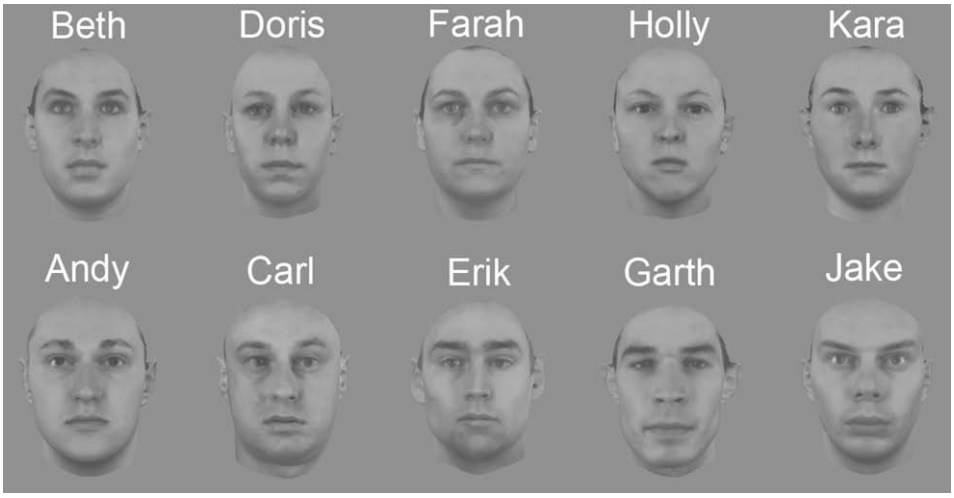

Fig. 11. Examples of the 10 unique faces (5 female; 5 male) selected for Experiment 4 along with the name assigned to each face.

5.1.4. Psychophysical procedure

5.1.4.1. Face training sessions. Before running the current experiment, participants were trained to recognize each of the 10 faces. This procedure involved two phases: (1) a study phase and (2) a test phase. Each study phase was immediately followed by a test phase. In the study phase, participants were randomly presented with one of the 10 faces with the associated name displayed below it. Each face and name was displayed until the participant felt that it had been properly "studied", at which time the participant indicated (via mouse click) readiness for the next face. This process was repeated five times for each of the 10 faces (i.e., 50 study trials). For the test phase, one of the 10 faces was randomly presented without the name label for 1000 ms, followed by a list of the 10 names which were displayed until the participant selected (via mouse click) the name thought to be associated with the presented face. This process was repeated five times for each of the 10 faces (i.e., 50 test trials). Stimulus duration was chosen according to that used by Näsänen (1999). Following completion of each study and test phase, participants were presented with the % correct for that session along with the number of errors (if any) for each individual face. For participants to successfully "pass" the test phase, they had to receive a score of 98% (which allowed for just one "click error"). The study and test phases were repeated until each participant received a recognition score of 98% or higher a minimum of three successive times. The study and test phases were spread out across multiple days.

5.1.5. Experiment sessions

Before the start of each experimental session, observers were run through a shortened study and test phase (one repetition of the study phase, followed by three repetitions of the test phase) in order to ensure that the observers could recognize the faces at the same level of accuracy as obtained in the initial face training sessions. If a participant had "failed" the pre-experiment study an test session, the experiment would have aborted and that participant would have been trained further on the face set (which never occurred for our participants).

If a given observer "passed" the pre-experiment (shortened) study and test session, the actual experimental session began. The general psychophysical task employed a 10AFC method of constant stimuli paradigm. Participants viewed the display at a distance of 27 cm in order to equate the visual angle of the faces to that of Experiments 1–3. For any given trial, observers were presented with a single stimulus interval (1000 ms) containing one of the PP_{FILT}-filtered faces (selected randomly), followed by a white noise mask (500 ms), followed by a response interval where the list of names from which the observers were to indicate (via mouse click) the name of the face preceding the white noise mask (the duration of the response interval was unlimited) was presented. Feedback was not provided, and all stimuli were viewed binocularly. All experimental sessions were grouped by central frequency of PP_{FILT}. All observers were allowed practice sessions to familiarize themselves with the task prior to engaging in the experimental sessions. Each session was repeated 4–6 times, with 150 trials per session (1800–2700 total trials). Threshold estimates for successful face recognition were assessed using "psignifit" (Wichmann & Hill, 2001a, 2001b), and were calculated for performance across PP_{FILT} bandwidth, for central face frequency. Threshold estimates were taken for the data averaged across the 10 faces for each observer.

5.1.6. Template-matching benchmark

For any given trial, the task of the psychophysical observers was to recognize a given PP_{FILT}-filtered face from a set of 10 possible non-PP_{FILT}-filtered faces. As in Experiment 1, performance could potentially be explained by a template-matching strategy. A correct response would be given according to which "internal" template face yielded the highest cross-correlation. Thus, template matcher performance was assessed by simulating the experiment described above, where, for each trial, each non-filtered template face was cross-correlated with each stimulus interval. The template matcher responded by selecting the highest cross-correlation between the "internal" templates and the particular PP_{FILT}-filtered face. Estimates of the template matcher's thresholds were made using the same procedures described in the psychophysical procedure section of the current experiment. The experiment was simulated four times in order to have one template matcher for each human observer that participated in the current experiment.

5.2. Results

The results from Experiment 4 are shown in Fig. 12. Since the four observers varied in their PP_{FILT} bandwidth thresholds, each observer's data are plotted individually. While the four observer's thresholds differed, there was a bias in favor of 9 cpf, with narrower PP_{FILT} bandwidths obtained when PP_{FILT} was centered on 9 cpf (with the exception of observer S.A. who showed this bias, but also had a low threshold for 3 cpf). On average, observers

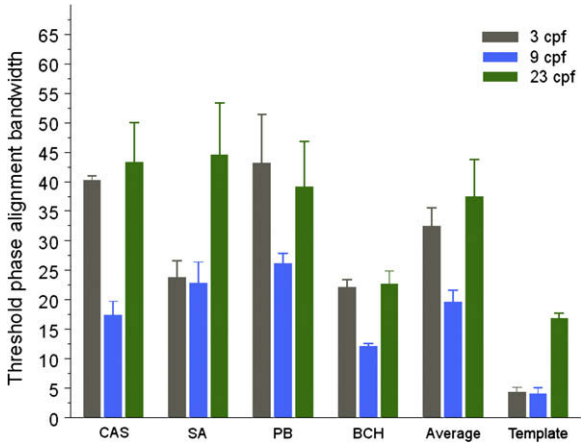

Fig. 12. Data from Experiment 4. On the ordinate is averaged number of aligned frequency threshold (i.e., PP_{FILT} bandwidth). On the abscissa is averaged data from each of the four subjects along with the averaged thresholds across all four subjects as well as the averaged ideal observer data. Each bar for each subject corresponds to a given central face frequency condition. Error bars are +1*SEM*, averaged across sessions for each observer. Refer to text for further details.

required a fewer number of phase alignments when PP_{FILT} was centered on 9 cpf when compared to 3 or 23 cpf. This observation was verified with a paired samples *t*-test assuming un-equal variances between the 9-cpf thresholds and the averaged thresholds for 3 and 23 cpf, $t_{(3)} = 4.77$, $p < .05$. This finding lends further support to the concept of a critical frequency for face recognition and provides a reasonable explanation for the differences observed in Experiment 1 of the current study and the phase randomization conditions in Näsänen's (1999) face recognition paradigm. It is also worth noting that the overall phase alignment thresholds were higher for face recognition (i.e., the current experiment), with an average of 30 phase-aligned frequencies, when compared to face discrimination (i.e., Experiment 1), with an average of 17 phase-aligned frequencies. Regarding the template matcher thresholds, Fig. 12 shows that equivalent PP_{FILT} bandwidth thresholds were produced when PP_{FILT} was centered on 3 or 9 cpf, with higher PP_{FILT} thresholds when PP_{FILT} was centered on 23 cpf. Here, the human observers performed the task less effectively than the template matcher as well as yielding a bias in favor a central face frequency of 9 cpf, a trend that was not observed with the template matcher, suggesting that when relying on phase alignment for face recognition, the template-matching strategy is likely not employed by human observers.

6. General discussion and conclusions

In Experiment 1 we observed a reduction in performance as a function of number of aligned frequencies, and not as a function of central frequency or octave bandwidth when compared with non-phase-randomized control data. In addition, the psychophysical discrimination thresholds showed that human performance demonstrated a more effective utilization of aligned frequencies than simple template matching would predict. Data from the second experiment suggested that the reduction in performance was likely due to the presence of non-aligned phases that interfered with the available phase-aligned components. In the third experiment, we demonstrated that the interference observed is likely to be band-limited as while the signal-to-noise ratio remained constant for the image, the task was rendered virtually impossible with increased spacing between the phase-aligned components. Taken together, the results from Experiments 1–3 stress the importance of *contiguous* critical bands for phase-aligned structure for discrimination between an average and unique faces, and likely observed regardless of where they are located within the face frequency spectrum. On the other hand, the final experiment, designed to test whether the trend in PP_{FILT} thresholds for face discrimination would be obtained in a face recognition paradigm, showed that (1) on average, more phase alignment was required to successfully recognize faces and (2) the central frequency at which PP_{FILT} was centered was critical. Specifically, human observers required less phase alignment when PP_{FILT} was centered at 9 cpf when compared to 3 or 23 cpf, providing further support for a critical frequency for face recognition. Additionally, human observers were found to be much less effective at making use of frequency alignment for face recognition, contrary to what was shown in Experiment 1 where face discrimination was investigated. Template-matching simulations were used to provide a reasonable benchmark for comparison to the behavioral data. The results for Experiments 1 and 4 argue that humans did not seem to employ such a strategy in our task. An alternative benchmark might have been to measure the identity strength of a face as a function of its distance from a mean face in a multidimensional face space (Leopold et al., 2001; Valentine, 1991) where the face at the origin of this space would approximate an average face (i.e., the prototypical face) and unique faces would be located along trajectories extending out from the prototype, with "uniqueness" increasing with distance from the prototype. The observers in our study had to discriminate identity from the prototype, thus a model which made use of the face space structure to make assessments might provide a more accurate benchmark for future studies.

The results from Experiments 1 and 4 clearly demonstrate substantial differences between the processes sub-serving face *discrimination* and those involved in face *recognition*. Specifically, face discrimination seems to require a fixed band of contiguous aligned phases, regardless of the central face frequency at which PP_{FILT} was centered, whereas face recognition, on average, required more phase alignment, with a narrower PP_{FILT} bandwidth centered at 9 cpf compared to 3 and 23 cpf.

The performance difference observed between face recognition and discrimination may speak to the nature of the stored representation of face. Given that the 9 cpf resulted in lower PP_{FILT} bandwidth thresholds (compared to 3 and 23 cpf in the same condition) suggests that a closer match is needed between the stored representation of the learned faces and their partially phase-randomized counterparts in order to name specific identities. In contrast, the absence of such a difference in the discrimination task suggests that the system may have been able to function without the bottleneck brought about by limits of multiple stored representations (i.e., no specific identity needed to be named on a trial-by-trial basis). In other words, face recognition may not be limited by perception, but perhaps by comparison to stored representations. Whether it is the stored representations themselves that are biased toward 9 cpf or the comparative process where this bias arises will be the subject of a future study.

The performance reduction observed in Experiment 1 was a function of the contiguous number of phase-aligned frequencies. We found no difference in performance between the three central frequencies we evaluated. Considered within our proposed framework, 16–18 contiguous frequencies define the minimum band of phase-aligned frequencies needed to adequately perform the face discrimination task. Yet, in Experiment 2 we found reliable performance with much smaller frequency bandwidths, once the phase-misaligned content outside the preserved band was removed. If the critical factor for performance in this task is the absolute number of phase-aligned frequencies, then it seems that our estimate from Experiment 1 is exaggerated. But if the critical factor is rather the *ratio* of phase-aligned to phase-misaligned content falling within a

band, then we would expect such robust performance on the linear-filtered images, even at very small bandwidths, because each band only contains phase-aligned components.

The notion that the *ratio* of phase-aligned to phase-misaligned face frequencies in a band as the determining factor of face discrimination performance does not necessarily mean that there are tuned mechanisms within the human visual system that extract such content from local regions of the frequency spectrum. This is because the ratio might simply reflect a global signal/noise limit averaged across all the frequencies in the image. We tested this in Experiment 3 by keeping the global signal-to-noise ratio of the image constant by holding the number of phase alignments constant at the critical number determined in Experiment 1 and varying how those aligned frequencies were distributed across the spectrum. The resulting deterioration in performance suggested that it was not the signal/noise ratio across all the image frequencies that was critical, but rather the signal/noise ratio across a contiguous band of frequencies. We define this as the critical band of phase alignments. Operationally, the concept is similar to that of a spatial channel. Traditionally, a spatial channel has been defined as a neural process tuned to a range of spatial frequencies (aligned with the receptive field). Here, we define a "phase-alignment channel" as a neural process that is tuned to phase alignments across a *contiguous* range of spatial frequencies (anywhere in the receptive field). We assume that these mechanisms are distributed equally and operate with equal efficiency across any region of the face frequency spectrum. Finally, there remains the issue of how well the findings reported here generalize to other object classes, an issue we are currently investigating.

6.1. Neural substrate

It is not possible using the current data to determine at which level in the visual pathway the operations that we suggest are crucial for face discrimination take place. Simple cells are sensitive to absolute phase relative to their receptive fields. Complex cells on the other hand are not tuned to absolute phase relative to their receptive field but are tuned to relative phase alignments across frequency in any position within the receptive field. For example, simple cells will respond similarly to both natural scene patches and scrambled phase versions of those patches, whereas complex cells respond better to natural scene patches as opposed to their scrambled phase versions, (Felsen, Touryan, Han, & Dan, 2005). The only differences between the scrambled phase and non-scrambled natural scene patches are in their phase spectra (i.e., both sets of stimuli possessed identical power spectra), suggesting that complex cells may represent an early stage of the type of phase alignment analysis that we have identified as important for the high-level task of face discrimination.

6.2. "Image-blocking" illusions

The identification of faces can be disrupted by coarse quantization (also known as blocking) of image luminances (Harmon & Julesz, 1973). It has been argued that this perceptual disruption cannot be simply explained by the increased amplitude of the high-frequency components introduced as a side-effect of the blocking (Morrone, Burr, & Ross, 1983) and that the phase spectrum plays an important, albeit presently undefined, role (Hayes, 1988). Within the context of the current investigation, the perceptual effects of image blocking rely on a disruption of local phase alignments; instead of randomizing phases as we have done here, image blocking disrupts them by artificially aligning them to the same value (i.e. the block boundary). In light of our current findings, it is therefore not surprising that the blocking effect has a perceptually profound disruption for face recognition.

Acknowledgments

This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) Grant: MT 108-18 to R.F.H. and a Doctoral fellowship from the Fonds de la Recherche en Santé du Québec to R.F.

References

Anderson, S. J., & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research, 25*, 1147–1154.
Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57*, 217–239.
Billock, V. A. (2000). Neural acclimation to 1/*f* spatial frequency spectra in natural images transduced by the human visual system. *Physica D, 137*, 379–391.
Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, 203*, 237–260.
Blanz, V. & Vetter, T. (1999). A morphable for the synthesis of 3D faces. Symposium on Interactive 3D Graphics. In *Proceedings of SIGGRAPH'99*, pp. 187–194.
Braje, W. L., Tjan, B. S., & Legge, G. E. (1995). Human efficiency for recognizing and detecting low-pass filtered objects. *Vision Research, 32*, 2955–2966.
Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics, 26*, 157–170.
Campbell, F. W., & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *Journal of Physiology, 181*, 576–593.
Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566.
Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York: Wiley. pp. 1–2.
Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics, 58*, 602–612.
De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*, 545–559.
Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biology, 3*, 1819–1828.
Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4*, 2379–2394.
Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception, 12*, 195–201.
Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural 24 processing of faces. *Perception, 34*, 77–86.
Goffaux, V., & Rossion, B. (2006). Faces are "spatial"–holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception & Performance, 32*, 1023–1039.
Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research, 39*, 3537–3560.
Graham, N. (1980). Spatial-frequency channels in human vision: Detecting edges without edge detectors. In C. Harris (Ed.), *Visual coding and adaptability* (pp. 215–252). Hillsdale, NJ: Erlbaum.
Hansen, B. C., & Essock, E. A. (2005). Influence of scale and orientation on the visual perception of natural scenes. *Visual Cognition, 12*, 1199–1234.
Hansen, B. C., & Hess, R. F. (2007). Structural sparseness and spatial phase alignment in natural scenes. *Journal of the Optical Society of America A, 24*, 1873–1885.
Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: Effects of twodimensional filtered noise. *Science, 180*, 1194–1197.
Hayes, A. (1988). Identification of two-tone images; some implications for high- and low-spatial-frequency processes in human vision. *Perception, 17*, 429–436.
Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception, 15*, 595–602.
Heckmann, T., & Schor, C. M. (1989). Is edge information for stereoacuity spatially channeled? *Vision Research, 29*, 593–607.
Hess, R. F., Wang, Y-Z., & Liu, C. H. (2006). The accessibility of spatial channels for stereo and motion. *Vision Research, 46*, 1318–1326.
Hess, R. F., Bex, P. J., Fredericksen, E. R., & Brady, N. (1998). Is human motion detection subserved by a single or multiple channel mechanism? *Vision Research, 38*, 259–266.
Ledgeway, T. (1996). How similar must the Fourier spectra of the frames of a random-dot kinematogram be to support motion perception? *Vision Research, 36*, 2489–2495.
Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4*, 89–94.
Maffei, L., & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vision Research, 13*, 1255–1267.
Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research, 42*, 1165–1184.
Morrone, M. C., Burr, D. C., & Ross, J. (1983). Added noise restores recognizability of coarse quantized images. *Nature, 305*, 226–228.
Morrone, M. C., & Owens, R. A. (1987). Feature detection from local energy. *Pattern Recognition Letters, 6*, 303–313.

Morrone, M. C. & Burr, D. C. (1988). Feature detection in human vision: A phase dependent energy model. *Proceedings of the Royal Society, London*, B, 235, 221–245.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research, 39*, 3824–3833.

Olzak, L. A., & Wickens, T. D. (1997). Discrimination of complex patterns: Orientation 25 information is integrated across spatial scale; spatial frequency and contrast information are not. *Perception, 26*, 1101–1120.

Owsley, C., Sekuler, R., & Boldt, C. (1981). Aging and low-contrast vision–face perception. *Investigative Ophthalmology & Visual Science, 21*, 362–365.

Phillips, G. C., & Wilson, H. R. (1984). Orientation bandwidths of spatial mechanisms measured by masking. *Journal of the Optical Society of America A, 1*, 226–232.

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing marilyn into maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8*, 107–113.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters, 73*, 814–817.

Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *NeuroImage, 32*, 1385–1394.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75*, 221–242.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369*, 395–397.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology A, 46*, 225–245.

Tieger, T., & Ganz, L. (1979). Recognition of faces in the presence of two-dimensional sinusoidal masks. *Perception & Psychophysics, 26*, 163–167.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3D objects in luminance noise. *Vision Research, 35*, 3053–3069.

Tolhurst, D. J., Tadmor, Y., & Chao, Tang (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics, 12*, 229–232.

Troje, N., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*, 1761–1771.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology A, 43*, 161–204.

van der Schaaf, A., & van Hateren, J. H. (1996). Modeling the power spectra of natural images: Statistics and Information. *Vision Research, 36*, 2759–2770.

White, M., & Li, J. (2006). Matching faces and expressions in pixelated and blurred photos. *American Journal of Psychology, 119*, 21–28.

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function. I: Fitting, sampling, and goodness of fit. *Perception and Psychophysics, 63*, 1293–1313.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function. II: Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics, 63*, 1314–1329.

Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research, 19*, 19–32.

Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research, 23*, 873–882.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*, 747–759.